The Application of Wavelet Transform and Transfer Learning for Gasoline Classification Using a Handheld Raman Spectrometer

Ting-Yu Huang, M.S.*, and Jorn Chi-Chung Yu, PhD
Department of Forensic Science, Sam Houston State University, Huntsville, TX 77340

MEMBER THE TEXAS STATE UNIVERSITY SYSTEM

ABSTRACT

Gasoline is one of the ignitable liquids (ILs) that has been most widely used as an accelerant to speed up fire escalation in arson cases. A rapid and accurate approach to detect gasoline allows investigators to recognize and preserve the evidence in time correctly and get a head start tracing the prime suspect during the investigation. Handheld Raman spectrometer combined with artificial intelligence (AI) is an attractive analytical platform that offers an automatic and field-deployable capability for precise gasoline and non-gasoline ILs classification. In this presentation, gasoline with three grades and seventeen different ILs were chosen for AI development using Raman spectra. Transfer learning based on a pre-trained deep convolutional neural network and wavelet transform analysis were performed to create a classifier to detect gasoline from various non-gasoline ILs.

INTRODUCTION

Gasoline is often encountered as physical evidence in a fire investigation. The main challenge for collecting and analyzing gasoline is that the evidence is subject to weathering and combustion. In addition, forensic analysis of gasoline from debris samples, according to ASTM E 1412-19 and E 1618-19, requires sample preparation procedures and instrumental analysis, indicating that the method is time-consuming and limited to laboratories.

Raman spectroscopy can provide information on chemical bonds and functional groups of organic molecules, and it is a well-known analytical method for molecule discrimination based on their vibrational frequencies. The portability, lower cost, and more straightforward analysis approach make it advantageous to real-time and on-site screening over other techniques. To increase the efficiency and automate the traditional spectrum matching method for identifying Raman spectra, deep learning algorithms are proposed in this study. Convolutional neural networks (CNNs) are the most dominant architecture in deep learning models and have held greatly increasing attention for data recognition and classification. The modular architectures utilize multiple layers to act as filters to extract features from unlabeled data and classify data without manual operation (1,2). However, training a new CNN from scratch depends heavily on huge data and is very computationally expensive. Therefore, repurposing a pre-trained CNN model, which is viewed as transfer learning, is a desirable alternative.

In this work, a handheld Raman spectrometer was used to collect gasoline and non-gasoline liquids spectra. After spectra acquisition, the continuous wavelet transform (CWT) was adopted to transform the spectra into images compatible with a pre-trained CNN to create a classifier. This research aims to propose a rapid, automated, and accurate approach for on-site gasoline detection and to demonstrate the potential for AI application in analyzing forensic evidence.

RESULTS AND DISCUSSION

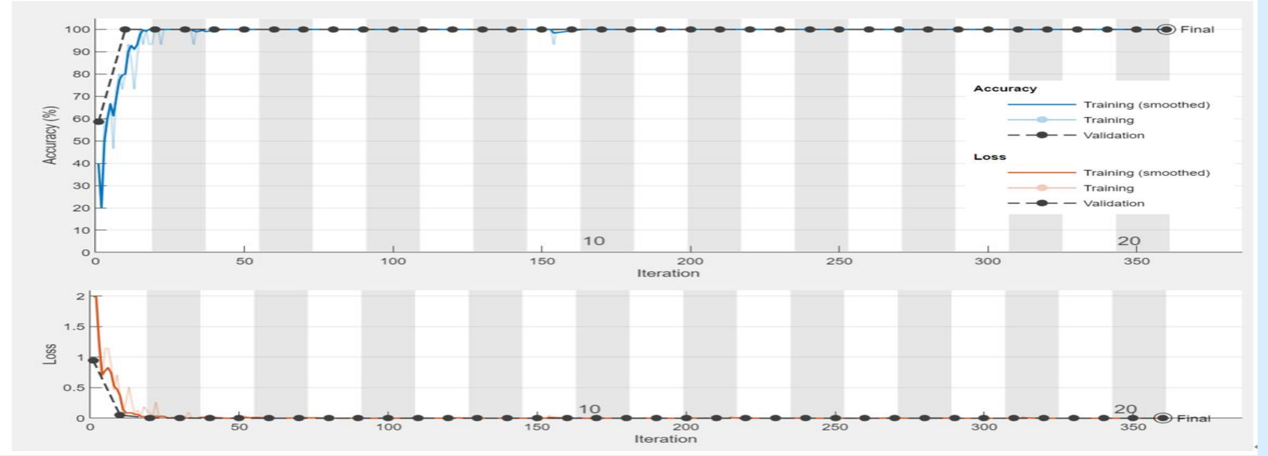
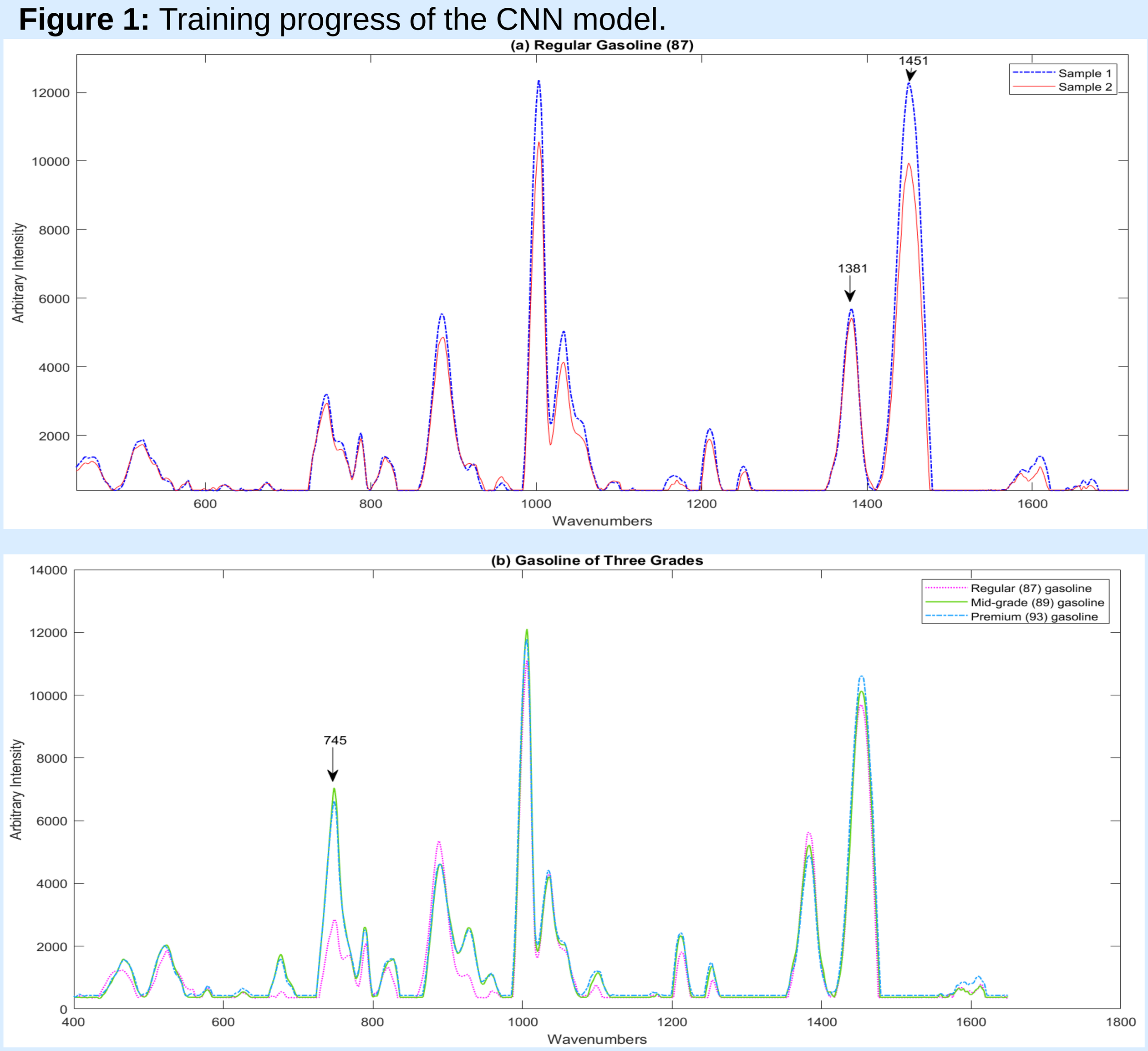


Figure 1: Training progress of the CNN model.



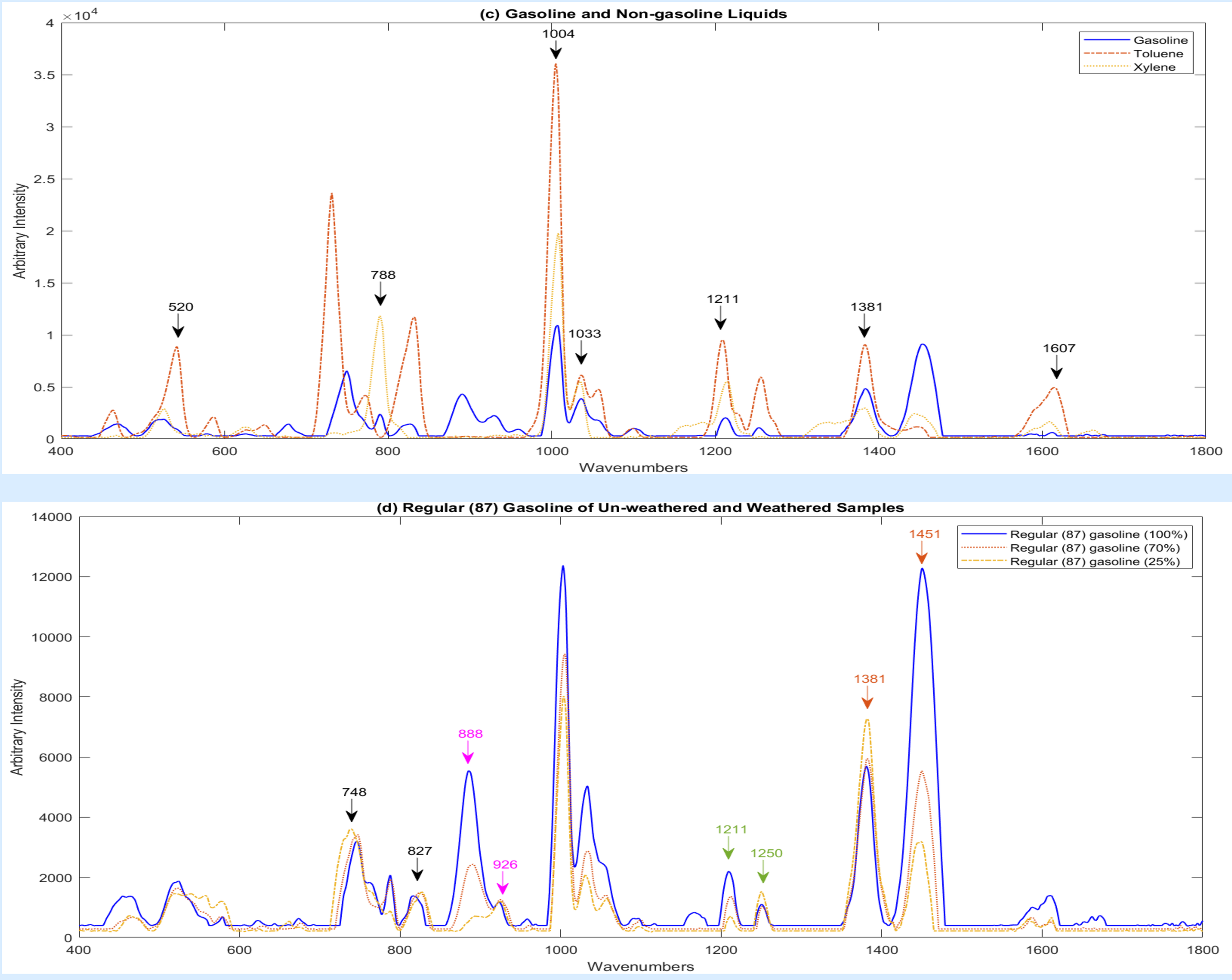


Figure 2: Examples of Raman spectra from the verification and evaluation dataset (a) Intensity variation of Raman scattering peaks in different sampling aliquots of gasoline (b) Spectral feature variation (745 cm^{-1}) of Raman scattering peaks between regular(87) gasoline, mid-grade(89) and premium(93) gasoline (c) Spectral feature similarities (520, 788, 1004, 1033, 1211, 1381, and 1607 cm^{-1}) of Raman scattering peaks between gasoline, toluene and xylene (d) Intensity, Raman shift (748, 827 cm^{-1}) and ratio (888,926, 1211,1250,1381,1451 cm^{-1}) variation of Raman scattering peaks in different levels of degraded gasoline.

Class	Predicted Gasoline	Predicted Non-gasoline	Predicted Gasoline	Predicted Non-gasoline	Predicted Gasoline	Predicted Non-gasoline
Gasoline	45	0	29	1	4	26
Non-gasoline Liquids	0	45	0	30	0	30
	Verification		Evaluation			
	100% Un-wearthered Gasoline		70% Wearthered Gasoline		25% Wearthered Gasoline	

Table 1: Classification results of the verification and the evaluation datasets.

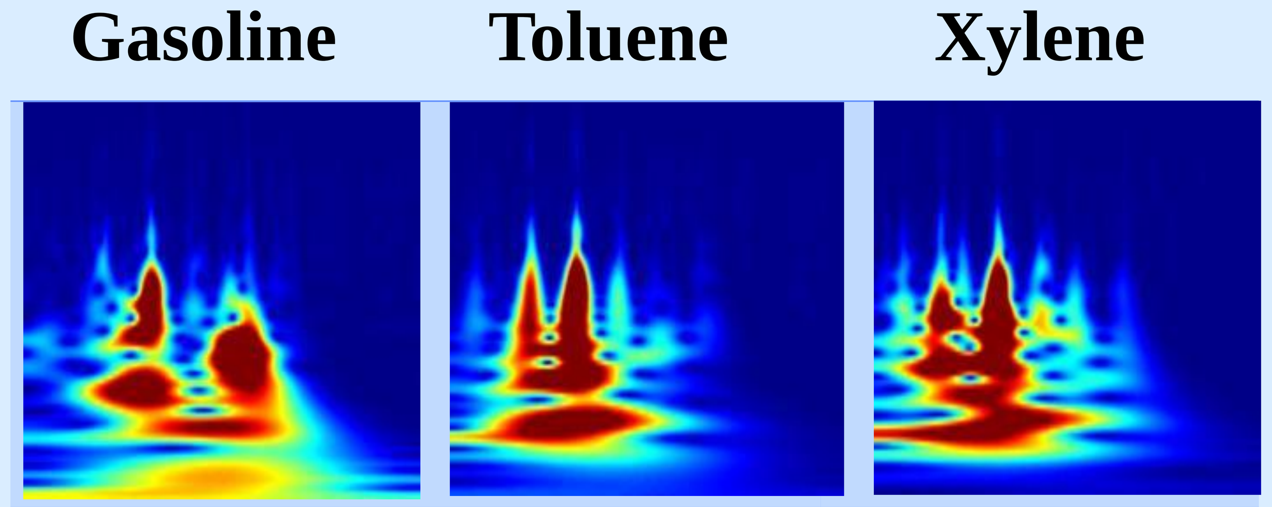


Figure 3: Examples of scalograms transformed from raw Raman spectra by CWT.

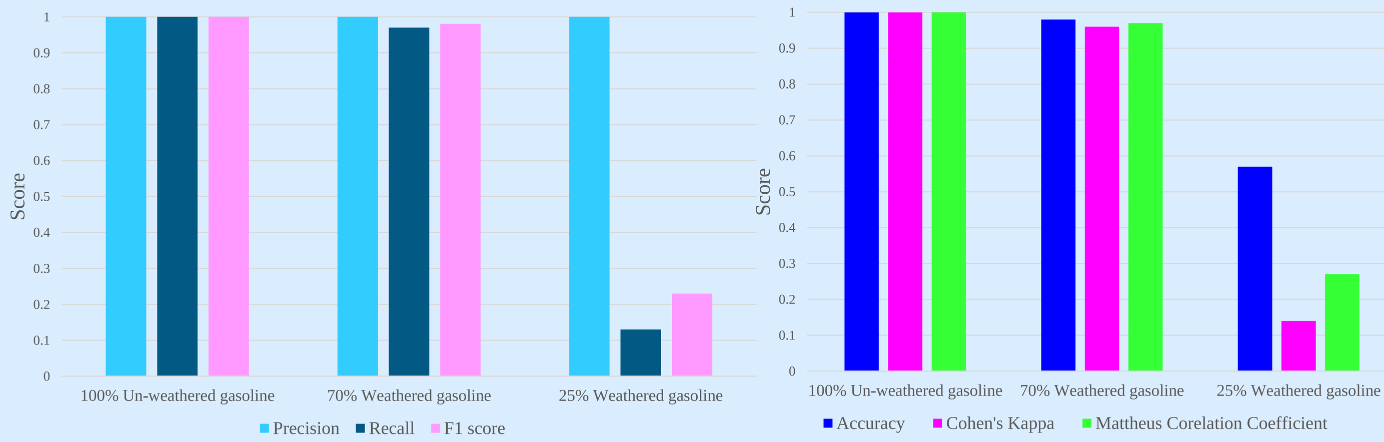


Figure 4: Performance indicators of the CNN model.

Model Training

The training dataset consisted of 180 gasoline scalograms and 170 non-gasoline scalograms. The training cycle was composed of 20 epochs and 360 iterations. The validation frequency was ten iterations. The validation accuracy reached 100% in the first epoch, and the training loss went to zero in the second epoch (Figure 1).

Model Verification

The verification dataset consisted of 45 gasoline scalograms and 45 non-gasoline scalograms. The evaluation dataset consisted of 60 gasoline scalograms and 60 non-gasoline scalograms. The spectral features of gasoline demonstrated intra-variations between different samples and between different octane numbers. The inter-similarities of characteristic peaks were also present in the spectra of gasoline and some non-gasoline liquids (Figure 2). The accuracy and the performance indicators of the CNN model all achieved 1 (100%), representing outstanding learning and generalization capabilities of interpreting complicated spectral information collected from gasoline.

The effect of gasoline sample degradation on the prediction probability was significant, with $F(2,102)=372.42$, $p=1.32\times10^{-47}$. The differences in the CNN model's classification performance was due to variation of peak intensity, Raman shift, and the ratio between levels of degradation. For gasoline samples degraded down to 70% weathered state, the CNN model could still maintain >90% classification performance (Figure 4).

MATERIALS AND METHODS

Sample Description and Raman Signal Measurement Three grades of gasoline, Regular (87), Mid-grade (89), and premium (93) from one fuel station in Houston, Texas, and seventeen non-gasoline liquids including kerosene, pure gum spirits turpentine, charcoal lighter, mineral spirit, VM&P naphtha, petroleum ether, ultra-grade performance vacuum oil, toluene, xylene, benzene, pentane, hexane, heptane, dodecane, ethanol, acetone, and nail polish from the market were obtained for this work. The handheld Raman with a 785 nm laser wavelength by HandyRamTM was used to acquire a total of 560 Raman spectra for training the model (350 spectra), verifying (90 spectra), and evaluating the performance of the model (120 spectra). The spectral range was recorded from 400 to 2300 cm^{-1} with 20 seconds of integration time and have been baseline-subtracted to correct the background signal for analysis.

Experimental Data Processing CWT was employed to create scalograms from Raman spectra, which were the absolute values of the CWT coefficients of the signals in the spectra plotted as a function of time and frequency. The parameters for CWT are depicted in Table 1. Each scalogram was generated in the format of an RGB image that was an array of size 224-by-224-by-3.

Deep CNN Training and Classification Methodology A pre-trained CNN, GoogLeNet, was adapted and fine-tuned to learn the new task of gasoline detection in this work. Fine-tuning consisted of replacing the final three layers (dropout, loss3-classifier, output) with new layers adapted to the new trained data and specifying training options as shown in Table 2. The scalograms from the training dataset were randomly divided into two parts, which were 80% (288 scalograms) and 20% (72 scalograms) for training and validating the model, respectively.

Types of Analysis	Parameter	Input
CWT Analysis	Signal Length	1000
	Sampling Frequency	Fs
	Voices Per Octave	12
Fine-tuning the Weights of the Pre-trained CNN Model	Mini Batch Size	15
	Max Epochs	20
	Initial Learn Rate	0.0001
	Validation Frequency	10

Table 2: Parameters of the CNN model.

CONCLUSIONS

- Novel approach integrated Raman spectrometry, wavelet analysis and deep learning for on-site gasoline detection.
- Wavelet analysis was successful in processing Raman spectra for feature extraction in AI development.
- No dependency on normalization and feature transformation by principal component analysis.
- Ceiling level of accuracy and high performance for non-degraded gasoline and degraded gasoline > 70% weathered state.

REFERENCES

1. Liu J, Osadchy M, Ashton L, Foster M, Solomon CJ, Gibson SJ. Deep convolutional neural networks for Raman spectrum recognition: A unified solution. The Analyst 2017;142(21):4067-74.
2. Rawat W, Wang Z. Deep convolutional neural networks for image classification: a comprehensive review. Neural Computation 2017;29(9):2352-449.

ACKNOWLEDGEMENTS

The authors thank Sam Houston State University staff for their insight and cooperation in collecting preliminary data.